University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2001 *Paper* 88

Mixture Hazards Models with Additive Random Effects Accounting for Treatment Effectiveness Lag Time

Ying Qing Chen* C. A. Rohde†
M.-C. Wang‡

*Division of Biostatistics, School of Public Health, University of California, Berkeley, yqchen@stat.berkeley.edu

†Department of Biostatistics, School of Hygiene and Public Health, Johns Hopkins University

‡Department of Biostatistics, School of Hygiene and Public Health, Johns Hopkins University

Mixture Hazards Models with Additive Random Effects Accounting for Treatment Effectiveness Lag Time

Ying Qing Chen, C. A. Rohde, and M.-C. Wang

Abstract

In many clinical trials to evaluate treatment efficacy, it is believed that there may exist latent treatment effectiveness lag times after which medical treatment procedure or chemical compound would be in full effect. In this article, semiparametric regression models are proposed and studied for estimating the treatment effect accounting for such latent lag times. The new models take advantage of the invariance property of the additive hazards model in marginalising over an additive latent variable; parameters in the models are thus easily estimated and interpreted, while the flexibility of not having to specify the baseline hazard function is preserved. Monte Carlo simulation studies demonstrate the appropriateness of the proposed semiparametric estimation procedure. The methodology is applied to data collected in a randomised clinical trial, which evaluates the efficacy of biodegradable carmustine polymers for treatment of recurrent brain tumours.

1 Introduction

In comparative randomised clinical trials, efficacy of a new treatment is often assessed by comparing survival data. However, some treatments may not always take effect as soon as they are applied, as illustrated by Pérez et al. (1997) in a report of a randomised, double-blind, placebo-controlled trial of a new antidepression treatment for patients with depressive symptoms. In such cases, researchers often believe that there may exist a so-called treatment effectiveness lag time before the treatment becomes fully effective (Wu et al., 1980; Gail, 1985; Lakatos, 1986; Zucker & Lakatos, 1990). A treatment effectiveness lag time is usually not observable, although certain biomarkers can be used to determine the termination of treatment effectiveness lag time artificially. If the treatment effectiveness lag times are ignored, the assumptions of the widely-used proportional hazards model (Cox, 1972) with constant proportionality are often inappropriate and hence cannot model the observation of slow onset of action correctly, unless some ad hoc time-dependent structure is included.

Hitherto, researchers have used the notion of a common fixed treatment effectiveness lag time for every individual or have tried to find ad hoc time-dependent lag functions for the proportional hazards model; see Self et al. (1988) and Zucker & Lakatos (1990). However, the treatment effectiveness lag times may well vary individually. In addition, since prior knowledge about the lag times is rarely available, a sensible lag function is usually unknown and not easily determined.

To account for the latent treatment effectiveness lag time and its heterogeneity among individuals, we introduce an unobservable latent variable, U, to represent such a lag time. Since some treatment effectiveness lag times are too long to allow the full onset of action, a mixture cure model (Farewell, 1982; Gray & Tsiatis, 1989; Laska & Meisner, 1992) will

be adopted for U. Furthermore, to identify the subject-dependent proportion of long-term treatment effectiveness lag times, appropriate regression models will be incorporated into the mixture cure model.

One way of modelling the treatment effect accounting for the latent U is to extend the proportional hazards model. For instance suppose that, given a treatment effectiveness lag time $U = u > 0$, the relative hazards ratio is assumed to be 1 before u and β after u. This provides the simplest version of the proportional hazards model with a so-called 'change point' as latent variable (Nguyen et al., 1984; Basu et al., 1988). After marginalisation over the latent variable, however, the clean multiplicative form is lost, leading to 'numerical and theoretical difficulties' in inference procedures and 'awkward interpretation' in parameters, as pointed out in Lin & Ying (1997).

In this article, we consider the additive hazards models with the treatment effectiveness lag time as latent change point. In §2, we will present the mixture model. The semiparametric inference procedure and its asymptotic properties are outlined in §3. Numerical studies are presented in §4; some concluding remarks and discussion are in §5; and technical proofs and computation are collected in several Appendices.

2 The mixture model

2.1 *Distribution of treatment effectiveness lag times*

Suppose that there are n independent participants in the study. For $i = 1, 2, \ldots, n$, the failure time and censoring time for individual i are T_i and C_i, respectively, and U_i is the latent treatment effectiveness lag time, after which the treatment is fully effective. The observed data consist of the triplets (X_i, Δ_i, Z_i). Here $X_i = \min(T_i, C_i)$ is the survival time,

and

$$\Delta_i = I(T_i \leqslant C_i) = \begin{cases} 1 & \text{if } T_i \leqslant C_i \\ 0 & \text{otherwise} \end{cases}$$

is the censoring indicator. Let $Z_i(t) = (W_i^{\mathrm{T}}(t), R_i^{\mathrm{T}}(t))^{\mathrm{T}}$ be the p-vector covariate, where superscript T denotes transpose. In a two-arm randomised clinical trial, for instance, W_i might be 1 if the participant is in the treatment group and 0 otherwise; and $R_i(t)$ contains the concomitant risk factors or confounding variables, for which the treatment effect needs to be adjusted. In addition, we assume that T_i and C_i are independent conditional on Z_i.

In practice, a proportion of population may never respond to the treatment (Pérez et al., 1997). Then their treatment effectiveness lag times are considered as 'long-term;' otherwise, they are 'short-term.' Denote by Υ_i the indicator of short-term U_i:

$$\Upsilon_i = \begin{cases} 1 & \text{if } U_i \text{ is short-term} \\ 0 & \text{if } U_i \text{ is long-term.} \end{cases}$$

Furthermore, let $\bar{F}_0(t;\tau) = 1 - F_0(t;\tau)$ be the conditional survival function for $\Upsilon_i = 1$, $i = 1, 2, \ldots, n$, where τ is parameter. Then the survival function of U_i, $\bar{G}_i(t) = 1 - G_i(t)$, $t \in [0, \infty)$, is assumed to be of the cure mixture form (Farewell, 1982):

$$\begin{aligned} \bar{G}_i(t;\tau) &= \mathrm{pr}(\Upsilon_i = 1) \times \mathrm{pr}(U_i \geqslant t | \Upsilon_i = 1) + \mathrm{pr}(\Upsilon_i = 0) \\ &= p_i \bar{F}_0(t;\tau) + (1 - p_i) \end{aligned} \tag{1}$$

for $i = 1, 2, \ldots, n$, where $p_i \in [0, 1]$. Here, the $F_0(t;\tau)$ might be exponential, Weibull or gamma. Seemingly, $G_i(t;\tau)$ in (1) is not a rigorously defined distribution function in theory, whenever $p_i < 1$. However, a truncation time, u_0 say, can be chosen to make G_i more mathematically concrete (Laska & Meisner, 1992; Tamura et al., 2000).

In (1), p_i can be associated with its corresponding covariate Z_i through appropriate regression models. For example, Farewell (1982) used the linear logistic regression model:

$$\log \frac{p_i(\alpha)}{1 - p_i(\alpha)} = \alpha^{\mathrm{T}} Z_i(0). \tag{2}$$

Other choices include probit, log-log and complementary log-log regression models (McCullagh & Nelder, 1989, p. 108). Furthermore, let $\phi_0 = (\tau_0^{\mathrm{T}}, \alpha_0^{\mathrm{T}})^{\mathrm{T}}$ be the true value of $\phi = (\tau^{\mathrm{T}}, \alpha^{\mathrm{T}})^{\mathrm{T}}$.

2.2 *Additive hazards models with latent lag time*

Denote by $\lambda(\cdot)$ the hazard function for the failure time, T. We propose the following model:

$$\lambda\{t|Z_i(t), U_i; \theta_0\} = \lambda_0(t) + \gamma_0^{\mathrm{T}} R_i(t) + I(U_i \leqslant t)\beta_0^{\mathrm{T}} W_i(t), \tag{3}$$

where $\theta_0 = (\beta_0^{\mathrm{T}}, \gamma_0^{\mathrm{T}})^{\mathrm{T}}$ is a p-vector parameter and $\lambda_0(t)$ is an unknown baseline hazard function. In (3), conditional on U_i, the hazard function of $Z_i(t)$ is $\lambda_0(t) + \gamma_0^{\mathrm{T}} R_i(t)$ before U_i and $\lambda_0(t) + \gamma_0^{\mathrm{T}} R_i(t) + \beta_0^{\mathrm{T}} W_i(t)$ after. The parameter β_0 therefore characterises the full effect of $W_i(t)$ after the lag time. This is often of the greatest interest, for example, when $W_i(t)$ is the treatment indicator.

Model (3) is a change point model which generalises the notion of fixed treatment effectiveness lag time (Zucker & Lakatos, 1990) by introducing heterogeneous U_i's. Jointly, models (1), (2) and (3) determine the probability with which the treatment is fully effective within a reasonable time range, along with the magnitude of its subject-specific effect.

Model (3) is also an additive hazards model with additive latent variables. Importantly, an additive hazards model yields a much simpler marginal model after the additive latent variables are integrated. For our proposed model, straightforward algebra shows that the marginalised model is

$$\lambda\{t|Z_i(t), \theta_0\} = \lambda_0(t) + \gamma_0^{\mathrm{T}} R_i(t) + \beta_0^{\mathrm{T}} W_i(t) H_i(t; \beta_0, \phi_0), \tag{4}$$

where

$$H_i(t; \beta_0, \phi_0) = \frac{\int_0^t e^{-\beta_0^{\mathrm{T}} \int_u^t W_i(s)ds} dG(u; \phi_0)}{\int_0^t e^{-\beta_0^{\mathrm{T}} \int_u^t W_i(s)ds} dG(u; \phi_0) + \bar{G}(t; \phi_0)}. \tag{5}$$

Although the above $H_i(t;\beta_0,\phi_0)$'s seem to be of complex algebraic form, they are intuitively meaningful under certain circumstances. For instance, when the treatment may reach full effect for every participant, i.e. $p_i \equiv 1$ in (1), then the $H_i(t;\beta_0,\phi_0)$'s have the following properties:

(i) $0 \leqslant H_i(t;\beta_0,\phi_0) \leqslant 1$;

(ii) $\lim_{t\to 0} H_i(t;\beta_0,\phi_0) = 0$, $\lim_{t\to\infty} H_i(t;\beta_0,\phi_0) = 1$;

(iii) $H_i(t;\beta_0,\phi_0)$ is nondecreasing.

These properties require $R_i(t)$ and $W_i(t)$ to be bounded, as well as the existence of an $i_0 \in \{1,2,\ldots,n\}$ such that $\mathrm{pr}(U_{i_0} \leqslant T_{i_0}) > 0$. The existence assumption is an identifiability condition that guarantees the estimability of β_0, which essentially requires that the treatment effectiveness lag time is not always longer than the failure time for every individual in study. Otherwise, $E\{I(U_i \leqslant T_i)\beta_0^{\mathrm{T}} W_i(t)\} \equiv 0$ for any i and then β_0 is not identifiable.

In fact, the $H_i(t;\beta_0,\phi_0)$'s are the lag functions that researchers have been seeking. It is also of interest that the $H_i(t;\beta_0,\phi_0)$'s have properties similar to those of cumulative distribution functions. When $W_i(t) \equiv 0$, $H_i(t)$ is exactly the distribution function of $G_i(t)$, although it does not have real impact on the model because of the zero $W_i(t)$. In §4, we study the nature of some special $H_i(t;\beta_0,\phi_0)$'s.

In general, identifiability of the parameters can be critical for models with arbitrary latent variables. However, this does not impose serious challenges upon model (3), as seen in the following theorem.

THEOREM 1. *If $W_i(t)$ and the parameter space $\mathcal{B}$ are bounded, then model (3) is identifiable if and only if ϕ in G_i is identifiable.*

To prove the above theorem, it is sufficient to show that two treatment effectiveness lag time distribution functions G_1 and G_2 are equal almost everywhere if

$$\int_0^\infty e^{-\beta_0^{\mathrm{T}} I(u\leqslant t)\int_u^t W_i(s)ds} dG_1(u;\phi_0) = \int_0^\infty e^{-\beta_0^{\mathrm{T}} I(u\leqslant t)\int_u^t W_i(s)ds} dG_2(u;\phi_0).$$

Its establishment is straightforward under the assumed conditions.

3 Inference Procedures and Asymptotic Properties

Since, as we have seen, the marginal model in (4) is in fact an additive hazards model with a special time-dependent coefficient, the semiparametric estimation procedure in Lin & Ying (1994) can be adapted. Let $N_i(t) = I(X_i \leqslant t, \Delta_i = 1)$ and $Y_i(t) = I(X_i \geqslant t)$, $i = 1, 2, \ldots, n$. Consider the filtration $\mathcal{F}_t$ defined by

$$\mathcal{F}_t = \sigma\{N_i(t), Y_i(t), Z_i(t); i = 1, 2, \ldots, n\}.$$

Define

$$M_i(t;\theta,\phi,\Lambda_0) = P_i(t;\theta,\phi) - \int_0^t Y_i(s)d\Lambda_0(s),$$

where $P_i(t;\theta,\phi) = N_i(t) - \int_0^t Y_i(s)\{\gamma^{\mathrm{T}} R_i(s) + \beta^{\mathrm{T}} W_i(s) H_i(s;\theta,\phi)\}ds$. Then $M_i(\cdot;\theta_0,\phi_0)$ are local square integrable martingales of $\mathcal{F}_t$. Therefore, similarly to the partial score equations for the proportional hazards model (Fleming & Harrington, 1991), the following estimating equations can be used to estimate (θ_0, ϕ_0):

$$\sum_{i=1}^n \int_0^\infty Q(t;\theta,\phi) J_i(t;\theta,\phi) dM_i(t;\theta,\phi,\Lambda_0) = 0, \tag{6}$$

where $Q(t;\theta_0,\phi_0)$ is a measurable weight function with respect to $\mathcal{F}_t$, and converges uniformly to a deterministic function $q(t;\theta_0,\phi_0)$. The $J_i(t;\theta,\phi)$ are smooth functions of the same dimension as (θ,ϕ), and are also predictable processes of t, $i = 1, 2 \ldots, n$.

The baseline hazard function in (6) is unknown, but a reasonable estimator of $\Lambda_0(t)$ of Breslow-type is

$$\hat{\Lambda}_0(t;\theta,\phi) = \int_0^t \left\{ \sum_{i=1}^n dP_i(s;\theta,\phi) \right\} \left\{ \sum_{i=1}^n Y_i(s) \right\}^{-1}, \tag{7}$$

as in Lin and Ying (1994). Thus we can use the following equations to estimate the parameters of interest:

$$\sum_{i=1}^n \int_0^\infty Q(t;\theta,\phi) J_i(t;\theta,\phi) d\hat{M}_i(t;\theta,\phi,\hat{\Lambda}_0) = 0, \tag{8}$$

where $\hat{M}_i(t;\theta,\phi,\hat{\Lambda}_0) = P_i(t;\theta,\phi) - \int_0^t Y_i(s) d\hat{\Lambda}_0(s;\theta,\phi)$. Denote the left-hand side of equation (8) by $\Gamma(\theta,\phi)$. Some algebraic manipulation shows that $\Gamma(\theta,\phi)$ is equal to

$$\Gamma(\theta,\phi) = \sum_{i=1}^n \int_0^\infty Q(t;\theta,\phi)\{J_i(t;\theta,\phi) - \bar{J}(t;\theta,\phi)\} dP_i(t;\theta,\phi), \tag{9}$$

where $\bar{J}(t;\theta,\phi) = \{\sum_{i=1}^n Y_i(t) J_i(t;\theta,\phi)\}\{\sum_{i=1}^n Y_i(t)\}^{-1}$. Consider the process $\Gamma(t;\theta,\phi) = \sum_{i=1}^n \int_0^t Q(s;\theta,\phi)\{J_i(s;\theta,\phi) - \bar{J}(s;\theta,\phi)\} dP_i(s;\theta,\phi)$. It is also true that $\Gamma(t;\theta_0,\phi_0)$ is an $\mathcal{F}_t$-martingale.

Let the solution of $\Gamma(\theta,\phi) = 0$ be $(\hat{\theta},\hat{\phi})$. The following theorem establishes the asymptotic properties of $(\hat{\theta},\hat{\phi})$.

THEOREM 2. *Suppose that there exists nonsingular $D(\theta_0,\phi_0)$ such that*

$$D(\theta_0,\phi_0) = \lim_{n\to\infty} -n^{-1} \begin{pmatrix} \partial\Gamma(\theta_0,\phi_0)/\partial\theta^{\mathrm{T}} \\ \partial\Gamma(\theta_0,\phi_0)/\partial\phi^{\mathrm{T}} \end{pmatrix}.$$

Furthermore, if all the partial derivatives are bounded and equicontinuous in a neighbourhood of (θ_0,ϕ_0), then under the regularity conditions 1-3 listed in the Appendix 1, $(\hat{\theta},\hat{\phi})$ is uniquely defined and

$$n^{1/2} \begin{pmatrix} \hat{\theta}-\theta_0 \\ \hat{\phi}-\phi_0 \end{pmatrix} \to N\{0, D^{-1}(\theta_0,\phi_0) V(\theta_0,\phi_0) D^{-1}(\theta_0,\phi_0)^{\mathrm{T}}\}, \tag{10}$$

in a neighborhood of (θ_0,ϕ_0), where $V(\theta_0,\phi_0) = V(t_0;\theta_0,\phi_0)$.

PROOF. See Appendix 1.

In fact, as pointed out in Lin & Ying (1995), if the equicontinuity condition is further satisfied on a compact region with whose interior contains the true parameter, then it is also possible to extend the above result globally.

For inference in practice the above asymptotic variance-covariance matrix is replaced by its consistent estimator $\hat{D}^{-1}(\hat{\theta},\hat{\phi})\hat{V}(\hat{\theta},\hat{\phi})\hat{D}^{-1}(\hat{\theta},\hat{\phi})^{\mathrm{T}}$. Furthermore, replacing the parameters of (θ,ϕ) with $(\hat{\theta},\hat{\phi})$ in (7) leads to a natural estimator of the cumulative baseline hazard function. As shown in Appendix 1, $n^{1/2}\{\hat{\Lambda}_0(t;\hat{\theta},\hat{\phi})-\Lambda_0(t)\}$ converges weakly to a zero-mean Gaussian process with covariance function $\Sigma(t_1,t_2)$.

Although the proposed estimating equations can be viewed as parallel to the partial score equations for the proportional hazards model, they are still ad hoc. However, by the techniques in Lai & Ying (1992) and Lin & Ying (1995), we can compute the semiparametric efficiency bound for the family of parametric submodels defined by

$$\lambda\{t|Z(t)\}=\lambda_0(t)+\gamma^{\mathrm{T}}R(t)+\beta^{\mathrm{T}}W(t)H(t;\beta,\phi)+\psi\eta(t),$$

where γ, β, ϕ and ψ are parameters and $\eta(\cdot)$ is a fixed function. As a result, the optimal estimating function for θ_0 is

$$\Gamma_{\mathrm{opt}}(\theta,\phi)=\sum_{i=1}^{n}\int_0^{\infty}\{\lambda_0(t)+\gamma^{\mathrm{T}}R_i(t)+\beta^{\mathrm{T}}W_i(t)H_i(t;\beta,\phi)\}^{-1}\{J_i^*(t;\theta,\phi)-\bar{J}^*(t;\theta,\phi,\lambda_0)\}dP_i(t;\theta,\phi),$$

where

$$J_i^*(t;\theta,\phi)=\begin{pmatrix} R_i(t) \\ W_i(t)H_i(t)+\beta^{\mathrm{T}}W_i(t)H'_{\beta}(t;\beta,\phi) \\ \beta^{\mathrm{T}}W_i(t)H'_{\phi}(t;\beta,\phi)\end{pmatrix},$$

$$\bar{J}^*(t;\theta,\phi,\lambda)=\frac{\sum_{i=1}^{n}Y_i(t)\{\lambda_0(t)+\gamma^{\mathrm{T}}R_i(t)+\beta^{\mathrm{T}}W_i(t)H_i(t;\beta,\phi)\}^{-1}J_i^*(t;\theta,\phi)}{\sum_{i=1}^{n}Y_i(t)\{\lambda_0(t)+\gamma^{\mathrm{T}}R_i(t)+\beta^{\mathrm{T}}W_i(t)H_i(t;\beta,\phi)\}^{-1}}.$$

However, it is difficult to use Γ_{opt} in practice, because the estimating functions themselves involve the baseline hazard function. Although adaptive procedures using special techniques such as sample-splitting (Lin & Ying, 1995) are available, the estimation of $\lambda_0(\cdot)$ is always a challenge, especially when sample size is small.

To implement the optimal estimating functions in practice, we can use similar versions instead for convenience. For example, one choice suggested in Lin & Ying (1994, 1995) is to use $\Gamma_{\rm opt}$ without including $\{\lambda_0(t)+\gamma^{\rm T}R_i(t)+\beta^{\rm T}W_i(t)H_i(t;\beta,\phi)\}^{-1}$, that is,

$$\Gamma^*(\theta,\phi)=\sum_{i=1}^{n}\int_0^{\infty}\{J_i^*(t;\theta,\phi)-\tilde{J}^*(t;\theta,\phi,\lambda_0)\}dP_i(t;\theta,\phi), \tag{11}$$

where

$$\tilde{J}^*(t;\theta,\phi,\lambda_0)=\frac{\sum_{i=1}^{n}Y_i(t)J_i^*(t;\theta,\phi)}{\sum_{i=1}^{n}Y_i(t)}.$$

When the ignored term is close to being a constant, Γ^* should not lead to much loss of efficiency.

4 Numerical Studies

4.1 *Examples of lag function* $H(t;\beta_0,\phi_0)$

As proposed in §2, the distribution of G is a cure mixture distribution. In this section, we work out lag functions for certain choices for G.

In G, $\bar{F}_0(t;\tau)$ is the conditional probability that the short-term lag time occurs after time t. Suppose that $F_0(t;\tau)$ corresponds to an exponential distribution

$$\bar{F}_0(t;\tau_0)=e^{-\tau_0 t}I(t\geqslant 0).$$

Furthermore, let $W(t)$ be a constant, W_0. Then, as seen in Appendix 2,

$$\begin{aligned}H(t;\beta_0,\phi_0)&=\frac{\tau_0 p(\alpha_0)}{\beta_0 W_0-\tau_0}(e^{-\tau_0 t}-e^{-\beta_0 W_0 t})\\&\times\left[\frac{\tau_0 p(\alpha_0)}{\beta_0 W_0-\tau_0}(e^{-\tau_0 t}-e^{-\beta_0 W_0 t})+p(\alpha_0)e^{-\tau_0 t}+\{1-p(\alpha_0)\}\right]^{-1},\end{aligned} \tag{12}$$

where ϕ_0 and $p(\alpha_0)$ are defined as in §2.1.

If instead treatment time is to be truncated, we can use, as in Gray & Tsiatis (1989) and Laska & Meisner (1992), the truncated exponential distribution of form

$$\frac{e^{-\tau_0 t} - e^{-\tau_0 u_0}}{1 - e^{-\tau_0 u_0}} I(0 \leqslant t \leqslant u_0).$$

Then

$$H(t; \beta_0, \phi_0) = \frac{\tau_0 p(\alpha_0)}{\beta_0 W_0 - \tau_0} \left(\frac{e^{-\tau_0 t} - e^{-\beta_0 W_0 t}}{1 - e^{-\tau_0 u_0}} \right)$$
$$\times \left[\frac{\tau_0 p(\alpha_0)}{\beta_0 W_0 - \tau_0} \left(\frac{e^{-\tau_0 t} - e^{-\beta_0 W_0 t}}{1 - e^{-\tau_0 u_0}} \right) + p(\alpha_0) \left(\frac{e^{-\tau_0 t} - e^{-\tau_0 u_0}}{1 - e^{-\tau_0 u_0}} \right) + \{1 - p(\alpha_0)\} \right]^{-1},$$

when $0 \leqslant t < u_0$; and $H(t; \beta_0, \phi_0) = 1$, otherwise. When u_0 goes to ∞, $H(t; \beta_0, \phi_0)$ tends to the form in (12).

To visualize $H(t; \beta_0, \phi_0)$, we plot it in various special situations. The lag functions are displayed for $p(\alpha_0)$ =1, 2/3, and 1/3 in Figs 1(a), 1(b) and 1(c), respectively, assuming that $W_i(t) \equiv 1$, $\beta_0 \equiv 1$, $\tau_0 = 1.01$, 1.5, 2.0 and $u_0 = \infty$. Since there is no finite truncation time for the treatment effectiveness lag times, the final lag functions are smooth. As shown in Fig 1(a), when all the treatment effectiveness lag times are short-term, then, marginally, the treatment will eventually reach full effectiveness. However, as shown in Figs 1(b) and 1(c), if there is any chance of a long-term lag time, the full effectiveness is not reachable. Instead, the treatment effect is washed out in the long run, although it may have some effect early on.

[Figure 1. about here]

The lag functions are displayed in Fig 2 for $u_0 = 5$, when everything else remains the same as in the setting of Fig 1. Since almost all the lag times are assumed to happen before u_0, it is not surprising to see the patterns similar to those in Fig 1 for $t < u_0$. However,

u_0 often serves as the termination of data collection in practice, we should not be able to observe anything informative after u_0. Nevertheless, as a demonstration, we show in Fig 2 that the treatment eventually reaches its full effectiveness if no lag time exists after u_0.

[Figure 2. about here]

Both figures show that, when τ_0 is bigger, the treatment effectiveness lag time becomes shorter, and then the mode of H is reached earlier, which means that the ultimate treatment effect is reached more quickly. Lag functions obtained from Weibull and gamma distributions can be found in Appendix 2 as well as the derivatives of H with respect to different parameters.

4.2 *Simulation studies*

Simulation studies have been conducted to study the performance of the estimation procedure proposed in §3. Two covariates are generated, R, which is continuous, following the Un(0,1) distribution, and W, which is 0 or 1 each with probability of 1/2, mimicking a treatment indicator. The baseline hazard function is chosen to be that of a Weibull distribution. Lag times are generated according to the mixture distribution in (1) with F_0 chosen to be exponential and $p(\alpha_0)$ to be constant. Then failure times are generated according to model (3), with $(\gamma_0, \beta_0) = (0,0)$, (0,1) and (1,0). Independent censoring times are generated from exponential distributions with different means to yield two censoring percentages of approximately 25% and 50%. Sample sizes are 100 and 200. Estimating functions in (11) are used for parameter estimation.

Simulation results are listed in Table 1. For each entry in the table, one thousand replicates are simulated for estimating the bias and empirical coverage probability. Here,

bias is defined as the difference between the sample mean of the estimates over the 1,000 simulated datasets and its respective true value; and 95% empirical coverage probability is the percentage of Wald-type 95% confidence intervals that include the true parameters. It is evident that the estimators are virtually unbiased and the nominal confidence intervals for the parameters have reasonable coverage probabilities.

[Table 1. about here]

4.3 *An example*

Our data come from a randomised placebo-controlled trial of the effectiveness of biodegradable carmustine polymers for the treatment of recurrent brain malignant gliomas (Brem et al., 1995). After the recurrent brain tumour was removed, a medicated or placebo polymer was placed to fill in the cavity. To reach a higher local drug concentration, the medicated polymers were supposed to release carmustine gradually over a two-to-three-week period following the placement (Tamargo et al., 1993; Brem et al., 1995). In the 27 medical centres of the trial, 222 patients were randomised to either the carmustine polymer treatment group (110 patients) or the placebo polymer group (112 patients). Their survival times measured in weeks, treatment assignment and prognostic factors can be found in Piantadosi (1997, pp. 496-509). Some exploratory analysis can be found in Brem et al. (1995) and Chen & Wang (2000).

In addition to the treatment indicator W, the age, R, is also considered as a covariate. Results from the proportional hazards model $\lambda(t|Z) = \lambda_0(t)\exp(\beta Z)$, the additive hazards model $\lambda(t|Z) = \lambda_0(t) + \beta Z$ and the proposed model (3) assuming F_0 to be exponential and the logistic regression model for response proportions, are listed in Table 2. The table shows that, although after adjusting for age the treatment effect does not appear significant in

either the proportional hazards model or the additive hazards model, it is significant if the treatment effectiveness lag time is taken into account. Given the presence of the lag time, the treatment will significantly decrease the hazard of the placebo group by 0.014, adjusting for age. The estimated average treatment effectiveness lag time is about 2.502 weeks for those who have short-term lag times. Its confidence interval does not contain 0, which implies the significant presence of such lag times. Furthermore, because of the non-significant α's, the proportion of short-term responders seems not to vary according to the treatment assignment or the participant's age.

[Table 2. about here]

5 Discussion

A more general mixture model for the treatment effectiveness lag time is

$$\bar{G}(t) = p\bar{F}_1(t) + (1-p)\bar{F}_2(t), \tag{13}$$

where $\bar{F}_1(t)$ is the survival function for a proportion p of the treatment effectiveness lag times and $\bar{F}_2(t)$ is the survival function for the rest. If $\bar{F}_2(t) \equiv 1$, then we obtain the cure mixture model. Furthermore, if the hazard function of F_1 is monotonically increasing, we should expect the hazard function of treatment effectiveness lag time to be initially increasing but decreasing later, as the patients with short-term treatment effectiveness lag times drop out of the risk set, leaving those with relatively longer treatment effectiveness lag times.

In the possible presence of treatment effectiveness lag time, the so-called 'intention-to-treat' principle (Sheiner & Rubin, 1995) may be applied to estimate the treatment effect. For example, if $W(t)$ is a binary treatment indicator in model (3), although the patient is a true 'control' before the treatment effectiveness lag time U, he or she should be counted

marginally as a member of treatment with probability of $H(t;\beta_0,\phi_0)$. As shown in Figure 1, when ϕ_0 increases, the treatment effectiveness lag time tends to be shorter and thus the treatment reaches its full effectiveness more quickly. So the intention-to-treat analysis of marginal treatment effect should approximate the subject-specific treatment effect.

So far as achieving simple inference procedure is concerned, the most critical aspect of the proposed model is the additivity of random effects. It is less critical whether or not the fixed effect of $R_i(t)$ is additive or multiplicative. Thus another class of change point hazards models with additive random effects is

$$\lambda\{t|Z_i(t),U_i;\theta_0\} = \lambda_0(t)e^{\gamma_0^{\mathrm{T}}R_i(t)} + I(U_i \leqslant t)\beta_0^{\mathrm{T}}W_i(t). \tag{14}$$

In contrast to the general additive-multiplicative hazards model in Lin & Ying (1995), it is not too difficult to find that model (14) is a parallel model but with random effects included in the additive component. Nevertheless, the marginalised model (14) should have the same H_i's as in (4). It is then straightforward to extend all the inference procedures and asymptotic results in previous sections to model (14).

Although the model proposed in this article has certain advantages, there are some important issues in its implementation. The first issue is inherited from the additive hazards model, namely, the parameter space is restricted by the magnitude of the baseline hazard function in order to obtain reasonable parameter estimates. One solution is to replace βZ with $\exp(\beta Z)$, but then the interpretation of β becomes cumbersome.

The second issue is inherited from the cure mixture model, namely, the potential identifiability problem with the parameters in the regression model of response proportions and the parameters in F_0. This has been a long-standing challenge for the cure mixture model. Theoretically, because of the potential over-parameterisation incurred by including the regression models for the p's, the nonsingularity of the matrix D is not secured, let alone its

positive-definiteness. As pointed out in Farewell (1998, pp. 1051-2), the estimator of these parameters here tend to have high correlation. This issue would be less critical if there were strong evidence to support the notion of existence of two heterogeneous populations. Otherwise, modelling just F_0 but ignoring the p's is adequate for detecting the possible existence of the lag time, estimating its average and deriving a sensible lag function.

Acknowledgement

The research of Y. Q. Chen was supported by the Hellman Family Faculty Fund. The research of M.-C. Wang was supported by the National Institutes of Health Grant R01 HD38209. The authors are grateful to the reviewers for helpful comments and the editor for invaluable editorial work.

Appendix 1

Asymptotics

We assume the following regularity conditions.

Condition 1. There exists a time $t_0 > 0$ such that $\lim_{n\to\infty} \sum_{i=1}^{n} Y_i(t_0) > 0$.

Condition 2. There exists an integrable function $v(t)$ such that, for any $t \in [0, t_0]$,

$$n^{-1} \sum_{i=1}^{n} Y_i(t)\{J_i(t;\theta_0,\phi_0) - \bar{J}(t;\theta_0,\phi_0)\}^{\otimes 2}\lambda_i(t|Z_i) - v(t) \to 0,$$

in probability, where $a^{\otimes 0} = 1$, $a^{\otimes 1} = a$ and $a^{\otimes 2} = aa^{\mathrm{T}}$.

Condition 3. For any $\epsilon > 0$,

$$n^{-1} \sum_{i=1}^{n} \int_0^{t_0} Y_i(s)\lambda(t|Z_i) \times \|J_i(t;\theta_0,\phi_0) - \bar{J}(t;\theta_0,\phi_0)\|^2 I\{n^{-1}\|J_i(t;\theta_0,\phi_0) - \bar{J}(t;\theta_0,\phi_0)\|^2 > \epsilon\} ds \to 0,$$

in probability, where $\|\cdot\|$ defines the Euclidean norm.

Weak convergence of $n^{-1/2}\Gamma(t;\theta_0,\phi_0)$. Since $n^{-1/2}\Gamma(t;\theta_0,\phi_0)$ is an $\mathcal{F}_t$-martingale process, the regularity conditions of 1-3 ensure that conditions (2.5.1) and (2.5.3) in Andersen et al. (1993, p. 83) are satisfied. Denote by $\mathfrak{D}[0,t_0]$ the space of the cadlag functions on $[0,t_0]$ endowed with the Skorohod topology. Then $n^{-1/2}\Gamma(t;\theta_0,\phi_0)$ converges weakly in $\mathfrak{D}[0,t_0]$ to a zero-mean Gaussian process with independent increments and variance function

$$V(t;\theta_0,\phi_0)=\lim_{n\to\infty} n^{-1}\sum_{i=1}^{n}\int_0^t Q(s;\theta_0,\phi_0)Y_i(s)\{J_i(t;\theta_0,\phi_0)-\bar{J}(t;\theta_0,\phi_0)\}^{\otimes 2}\lambda_i(t|Z_i)ds.$$

In fact, more delicate arguments can be used as in Ying (1993) to extend t_0 to ∞.

Asymptotics of $n^{1/2}(\hat{\theta}^{\mathrm{T}}-\theta_0^{\mathrm{T}},\hat{\phi}^{\mathrm{T}}-\phi_0^{\mathrm{T}})^{\mathrm{T}}$. It is not difficult to see that

$$\begin{aligned}
&-n^{-1}\begin{pmatrix}\partial\Gamma(t_0;\theta_0,\phi_0)/\partial\theta_0\\ \partial\Gamma(t_0;\theta_0,\phi_0)/\partial\phi_0\end{pmatrix}\\
=\;& n^{-1}\sum_{i=1}^{n}\int_0^{t_0} Y_i(t)\{J_i(t;\theta_0,\phi_0)-\bar{J}(t;\theta_0,\phi_0)\}\{\gamma_0^{\mathrm{T}}R_i(s)+\beta_0^{\mathrm{T}}W_i(s)H_i(s;\theta_0,\phi_0)\}'dt\\
&-n^{-1}\sum_{i=1}^{n}\int_0^{t_0}\{J_i(t;\theta_0,\phi_0)-\bar{J}^*(t;\theta_0,\phi_0)\}'dP_i(t;\theta_0,\phi_0),
\end{aligned}$$

where $(\cdot)'$ denotes the derivative of $(\cdot)$ with respect to (θ,ϕ). The second term on the right-hand side in the above equation is an average of martingale integrals and therefore converges in probability to zero. Under the conditions in Theorem 2, the first term converges to D. Therefore, $-n\Gamma'(t_0;\theta_0,\phi_0)\to D$ in probability. In fact, if we follow similar arguments to those in Lin & Ying (1995), it also holds that $(\hat{\theta}^{\mathrm{T}},\hat{\phi}^{\mathrm{T}})^{\mathrm{T}}\to(\theta_0^{\mathrm{T}},\phi_0^{\mathrm{T}})^{\mathrm{T}}$ in probability. Furthermore, by the Taylor expansion of $\Gamma(t_0;\hat{\theta},\hat{\phi})$ at (θ_0,ϕ_0), we know that

$$n^{1/2}\begin{pmatrix}\hat{\theta}-\theta_0\\ \hat{\phi}-\phi_0\end{pmatrix}=\left\{-n^{-1}\begin{pmatrix}\partial\Gamma(t_0;\theta^*,\phi^*)/\partial\theta^{\mathrm{T}}\\ \partial\Gamma(t_0;\theta^*,\phi^*)/\partial\phi^{\mathrm{T}}\end{pmatrix}\right\}^{-1}n^{-1/2}\Gamma(t_0;\theta_0,\phi_0),$$

where (θ^*,ϕ^*) is on the line segment between $(\hat{\theta},\hat{\phi})$ and (θ_0,ϕ_0). Then, with the equicontinuity assumptions and the theorems in Goffman (1965, Theorems 4.1 and 4.2) as used in

Lin and Ying (1995), the local uniqueness and the asymptotic normality of estimators are established. A straightforward variance calculation leads to the results in Theorem 2.

Asymptotics of $n^{1/2}\{\hat{\Lambda}_0(t;\hat{\theta},\hat{\phi}) - \Lambda_0(t)\}$. Using the decomposition technique outlined in Fleming & Harrington (1991, p. 300), we have that

$$\underbrace{n^{1/2}\{\hat{\Lambda}_0(t;\hat{\theta},\hat{\phi}) - \hat{\Lambda}_0(t;\theta_0,\phi_0)\}}_{\text{I}} + \underbrace{n^{1/2}\{\hat{\Lambda}_0(t;\theta_0,\phi_0) - \hat{\Lambda}_0^*(t)\}}_{\text{II}} + \underbrace{n^{1/2}\{\Lambda_0^*(t) - \Lambda_0(t)\}}_{\text{III}}, \quad \text{(A1)}$$

where $\Lambda_0^*(t) = \int_0^t I\{\sum_{i=1}^n Y_i(s) > 0\}\lambda_0(s)ds$. From the Taylor expansion and results in Theorem 2.2 in Lin & Ying (1995), term (I) is

$$\lim_{n\to\infty} K_1(t)D^{-1}n^{-1/2}\Gamma(t;\theta_0,\phi_0) + o_p(1),$$

uniformly for $t \in [0, t_0]$. In addition, term (II) is $o_p(1)$ by the Lenglart inequality (Andersen et al., 1993). It is also straightforward to see that term (III) is asymptotically ignorable. Therefore, by the multivariate martingale central limit theorem, the asymptotic properties of $\hat{\Lambda}$ are established. Straightforward variance calculation shows that $\Sigma(t_1, t_2)$ is the limit of

$$\begin{aligned} & \int_0^{\min(t_1,t_2)} \frac{n}{\sum_{i=1}^n Y_i(t)} \frac{\sum_{i=1}^n dN_i(t)}{\sum_{i=1}^n Y_i(t)} \\ + \quad & K_1^{\mathrm{T}}(t_2)D^{-1}V(D^{-1})^{\mathrm{T}}K_1(t_1) - K_1^{\mathrm{T}}(t_2)D^{-1}K_1(t_1) - K_1^{\mathrm{T}}(t_1)D^{-1}K_2(t_2), \end{aligned}$$

where

$$\begin{aligned} K_1(t;\theta_0,\phi_0) &= \int_0^t \left[\sum_{i=1}^n Y_i(s)\{\gamma_0^{\mathrm{T}}R_i(s) + \beta_0^{\mathrm{T}}W_i(s)H_i(s;\theta_0,\phi_0)\}'\right] \left\{\sum_{i=1}^n Y_i(s)\right\}^{-1} ds, \\ K_2(t;\theta_0,\phi_0) &= \int_0^t \left(\left[\sum_{i=1}^n Y_i(s)\{\gamma_0^{\mathrm{T}}R_i(s) + \beta_0^{\mathrm{T}}W_i(s)H_i(s;\theta_0,\phi_0)\}J_i(s;\theta_0,\phi_0)\right] \left\{\sum_{i=1}^n Y_i(s)\right\} \right. \\ &\quad \left. - \left[\sum_{i=1}^n Y_i(s)\{\gamma_0^{\mathrm{T}}R_i(s) + \beta_0^{\mathrm{T}}W_i(s)H_i(s;\theta_0,\phi_0)\}\right] \left\{\sum_{i=1}^n Y_i(s)J_i(s;\theta_0,\phi_0)\right\} \right) \\ &\quad \times \left\{\sum_{i=1}^n Y_i(s)\right\}^{-2} ds. \end{aligned}$$

Appendix 2

Examples of lag functions

Exponential lag time: $f(t) = \tau_0 e^{-\tau_0 t}$.

$$
\begin{aligned}
H &= \left\{ \frac{\tau_0 p(\alpha_0)}{\beta_0 W_0 - \tau_0} \left(e^{-\tau_0 t} - e^{-\beta_0 W_0 t}\right) \right\} \left\{ \frac{\tau_0 p(\alpha_0)}{\beta_0 W_0 - \tau_0} \left(e^{-\tau_0 t} - e^{-\beta_0 W_0 t}\right) + p(\alpha_0) e^{-\tau_0 u} + 1 - p(\alpha_0) \right\}^{-1} \\
H'_{\beta_0} &= \left\{ \tau_0 W_0 p(\alpha_0) \left((1 - \tau_0 t + \beta_0 W_0 t) \left[\{1 - p(\alpha_0)\} e^{-\beta_0 W_0 t} + p(\alpha_0) e^{-t(\tau_0 + \beta_0 W_0)} \right] \right.\right. \\
&\quad \left.\left. - \left\{ 1 + p(\alpha_0) e^{-\tau_0 t} - p(\alpha_0) \right\} e^{-\tau_0 t} \right) \right\} \\
&\quad \times \left\{ \tau_0 p(\alpha_0) e^{-\beta_0 W_0 t} - p(\alpha_0) e^{-\tau_0 t} \beta_0 W_0 - \beta_0 W_0 + p(\alpha_0) \beta_0 W_0 + \tau_0 - \tau_0 p(\alpha_0) \right\}^{-2} \\
H'_{\tau_0} &= p(\alpha_0) \left(\left[\{1 - p(\alpha_0)\} \left(\tau_0^2 t - \tau_0 \beta_0 W_0 t + \beta_0 W_0 \right) + p(\alpha_0) \beta_0 W_0 e^{-\tau_0 t} \right] e^{-\tau_0 t} \right. \\
&\quad \left. - \{1 - p(\alpha_0)\} \beta_0 W_0 e^{-\beta_0 W_0 t} + p(\alpha_0) \left(\tau_0^2 t - \tau_0 \beta_0 W_0 t - \beta_0 W_0 \right) e^{-t(\tau_0 + \beta_0 W_0)} \right) \\
&\quad \times \left\{ \tau_0 p(\alpha_0) e^{-\beta_0 W_0 t} - p(\alpha_0) e^{-\tau_0 t} \beta_0 W_0 - \beta_0 W_0 + \tau_0 + p(\alpha_0) \beta_0 W_0 - \tau_0 p(\alpha_0) \right\}^{-2} \\
H'_p &= \left(\beta_0 W_0 - \tau_0 \right) \left(e^{-\tau_0 t} - e^{-\beta_0 W_0 t} \right) \tau_0 \\
&\quad \times \left\{ \tau_0 p(\alpha_0) e^{-\beta_0 W_0 t} - p(\alpha_0) e^{-\tau_0 t} \beta_0 W_0 - \beta_0 W_0 + \tau_0 + p(\alpha_0) \beta_0 W_0 - \tau_0 p(\alpha_0) \right\}^{-2}.
\end{aligned}
$$

Weibull lag time: $f(t) = \tau_{10}^{\tau_{20}} \tau_{20} t^{\tau_{20}-1} e^{-(\tau_{10} t)^{\tau_{20}}}$.

$$
\begin{aligned}
H &= p(\alpha_0) \tau_{01}^{\tau_{02}} \tau_{02} e^{-\beta_0 W_0 t} \int_0^t e^{\beta_0 W_0 u - \tau_{01}^{\tau_{02}} u^{\tau_{02}}} u^{\tau_{02}-1} du \\
H'_{\beta_0} &= p(\alpha_0) \tau_{01}^{\tau_{02}} \tau_{02} W_0 e^{-\beta_0 W_0 t} \left(\int_0^t e^{\beta_0 W_0 u - (\tau_{01} u)^{\tau_{02}}} u^{\tau_{02}} du - t \int_0^t e^{\beta_0 W_0 u - (\tau_{01} u)^{\tau_{02}}} u^{\tau_{02}-1} du \right) \\
H'_{\tau_{01}} &= p(\alpha_0) \tau_{01}^{\tau_{02}-1} \tau_{02}^2 e^{-\beta_0 W_0 t} \left(\int_0^t e^{\beta_0 W_0 u - (\tau_{01} u)^{\tau_{02}}} u^{\tau_{02}-1} du - \tau_{01}^{2\tau_{02}} \int_0^t e^{\beta_0 W_0 u - (\tau_{01} u)^{\tau_{02}}} u^{2\tau_{02}-1} du \right) \\
H'_{\tau_{02}} &= p(\alpha_0) \tau_{01}^{\tau_{02}} e^{-\beta_0 W_0 t} \left[\{1 + (\log \tau_{01}) \tau_{02}\} \int_0^t e^{\beta_0 W_0 u - (\tau_{01} u)^{\tau_{02}}} u^{\tau_{02}-1} du \right. \\
&\quad \left. + \tau_{02} \int_0^t e^{\beta_0 W_0 u - (\tau_{01} u)^{\tau_{02}}} \left(-\tau_{01}^{\tau_{02}} \log \tau_{01} u^{\tau_{02}} - \tau_{01}^{\tau_{02}} u^{\tau_{02}} \log u + \log u \right) u^{\tau_{02}-1} du \right] \\
H'_p &= \tau_{01}^{\tau_{02}} \tau_{02} e^{-\beta_0 W_0 t} \int_0^t e^{\beta_0 u W_0 - (\tau_{01} u)^{\tau_{02}}} u^{\tau_{02}-1} du.
\end{aligned}
$$

Gamma lag time: $f(t) = \Gamma^{-1}(\tau_{02})\,\tau_{01}^{\tau_{02}} t^{\tau_{02}-1} e^{-\tau_{01} t}$.

$$
\begin{aligned}
H &= \frac{\tau_{01}^{\tau_{02}} p(\alpha_0) e^{-\beta_0 W_0 t}}{\Gamma(\tau_{02})} \int_0^t u^{\tau_{02}-1} e^{(\beta_0 W_0 - \tau_{01})u} u^{\tau_{02}-1} du \\
H'_{\beta_0} &= \frac{-\tau_{01}^{\tau_{02}} p(\alpha_0) W_0 e^{-\beta_0 W_0 t}}{\Gamma(\tau_{02})} \left(t \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2(\tau_{02}-1)}\, du - \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2\tau_{02}-1} du \right) \\
H'_{\tau_{01}} &= \frac{\tau_{01}^{\tau_{02}-1} p(\alpha_0) e^{-\beta_0 W_0 t}}{\Gamma(\tau_{02})} \left(\tau_{02} \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2(\tau_{02}-1)} du - \tau_{01} \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2\tau_{02}-1} du \right) \\
H'_{\tau_{02}} &= -\frac{\Gamma'(\tau_{02})\,\tau_{01}^{\tau_{02}} p(\alpha_0) e^{-\beta_0 W_0 t}}{\Gamma^2(\tau_{02})} \int_0^t u^{\tau_{02}-1} e^{(\beta_0 W_0 - \tau_{01})u} u^{\tau_{02}-1} du \\
&\quad + \frac{\tau_{01}^{\tau_{02}} p(\alpha_0) e^{-\beta_0 W_0 t}}{\Gamma(\tau_{02})} \left(\log \tau_{01} \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2\tau_{02}-2} du + 2 \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2\tau_{02}-2} \log u\, du \right) \\
H'_p &= \frac{\tau_{01}^{\tau_{02}} e^{-\beta_0 W_0 t}}{\Gamma(\tau_{02})} \int_0^t e^{(\beta_0 W_0 - \tau_{01})u} u^{2\tau_{02}-2} du.
\end{aligned}
$$

References

Andersen, P. K., Borgan, Ø., Gill, R. D. and Keiding, N. (1993). *Statistical Models based on Counting Processes.* New York: Springer-Verlag.

Basu, A. P., Ghosh, J. K. & Joshi, S. N. (1988). On estimating change point in a failure rate. *Statist. Decis.* **4**, 239-52.

Brem, H., Piantadosi, S., Burger, P. C., Walker, M., Selker, R., Vick, N. A., Black, K., Sisiti, M., Brem, S., Mohr, G., Muller, P. & Morawetz, R. (1995). Placebo-controlled trials of safety and efficacy of intraoperative controlled delivery by biodegradable polymers of chemotherapy of recurrent gliomas. *Lancet* **345**, 1008-12.

Chen, Y. Q. & Wang, M.-C. (2000). Analysis of accelerated hazards model. *J. Am. Statist. Assoc.* **95**, 608-18.

Cox, D. R. (1972). Regression models and life tables (with Discussion). *J. R. Statist. Soc.* B **34**, 187-220.

FAREWELL, V. T. (1982). The use of mixture models for the analysis of survival data with long-term survivors. *Biometrics* **38**, 1041-6.

FAREWELL, V. T. (1998). Cure models. In *Encyclopedia of Biostatistics*, Ed. P. Armitage and T. Colton, pp. 1051-3. Chichester: John Wiley.

FLEMING, T. R. & HARRINGTON, D. P. (1991). *Counting Processes and Survival Analysis.* New York: John Wiley.

GAIL, M. H. (1985). Applicability of sample size calculations based on a comparison of proportions for use with the log rank test. *Contr. Clin. Trials* **6**, 112-9.

GOFFMAN, C. (1965). *Calculus of Several Variables.* New York: Harper and Row.

GRAY, R. J. & TSIATIS, A. A. (1989). A linear rank test for use when the main interest is in differences in cure rates. *Biometrics* **45**, 899-904.

LAKATOS, E. (1986). Sample size determination in clinical trials with time-dependent rates of losses and noncompliance. *Contr. Clin. Trials* **7**, 189-99.

LAI, T. L. & YING, Z. (1992). Asymptotically efficient estimation in censored and truncated regression models. *Statist. Sinica* **2**, 17-46.

LASKA, E. M. & MEISNER, M. J. (1992). Nonparametric estimation and testing in a cure model. *Biometrics* **48**, 1223-34.

LIN, D. Y. & YING, Z. (1994). Semiparametric analysis of the additive risk model. *Biometrika* **81**, 61-71.

LIN, D. Y. & YING, Z. (1995). Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *Ann. Statist.* **5**, 1712-34.

Lin, D. Y. & Ying, Z. (1997). Additive hazards regression models for survival data. In *Lecture Notes in Statistics* **123**, Ed. D. Y. Lin and T. R. Fleming, pp. 185-98. New York: Springer-Verlag.

McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. New York: Chapman and Hall.

Nguyen, H. T., Rogers, G. S. & Walker, E. A. (1984). Estimation in change-point hazard rate models. *Biometrika* **71**, 299-304.

Pérez, V., Gilaberte, I., Faries, D., Alvarez, E., & Artigas, F. (1997). Randomised, double-blind, placebo-controlled trial of pindolol in combination with fluxetine antidepressant treatment. *Lancet* **349**, 1594-7.

Piantadosi, S. (1997). *Clinical Trials: A Methodologic Perspective.* New York: John Wiley.

Self, S., Prentice, R., Iverson, D., Henderson, M., Thompson, D., Byar, D., Insull, W., Gorbach, S. I., Clifford, C., Goldman, S., Urban, N., Sheppard, L. & Greenwald, P. (1988). Statistical design of the women's health trial. *Contr. Clin. Trials* **9**, 119-36.

Sheiner, L. B. & Rubin, D. B. (1995). Intention-to-treat analysis and the goals of clinical trials, *Clin. Pharmacol. Therap.* **57**, 6-15.

Tamargo, R. J., Myseros, J. S., Epstein, J. I., Yang, M. B., Chasin, M. & Brem, H. (1993). Interstitial chemotherapy of the 9L gliosarcoma: controlled release polymers for drug delivery in the brain. *Cancer Res.* **53**, 329-33.

Tamura, R. N., Faries, D. E. & Feng, J. (2000). Comparing time to onset of response in antidepressant clinical trials using the cure model and the Cramer-von Mises test.

Statist. Med. **19**, 2169-84.

Wu, M., Fisher, M., & DeMets, D. (1980). Sample sizes for long-term medical trial with time-dependent dropout and event rates. *Contr. Clin. Trials* **1**, 109-21.

Ying, Z. (1993). A large sample study of rank estimation for censored regression data. *Ann. Statist.* **21**, 76-99.

Zucker, D. M. & Lakatos, E. (1990). Weighted log rank type statistics for comparing survival curves when there is a time lag in the effectiveness of treatment. *Biometrika* **77**, 853-64.

Table 1: Summary of simulation studies. Each entry is the estimated bias, with the associated 95% empirical coverage probability in brackets.

		$(\gamma_0, \beta_0) = (0, 0)$		$(\gamma_0, \beta_0) = (0, 1)$	
n	%	R	W	R	W
100	25%	-0.056(0.950)	-0.038(0.962)	-0.005(0.959)	0.042(0.941)
	50%	0.078(0.927)	0.021(0.955)	-0.058(0.941)	0.026(0.930)
200	25%	0.008(0.949)	-0.029(0.944)	-0.006(0.961)	-0.021(0.960)
	50%	-0.026(0.951)	-0.008(0.941)	-0.025(0.946)	-0.015(0.964)

		$(\gamma_0, \beta_0) = (1, 0)$	
n	%	R	W
100	25%	-0.056(0.948)	0.024(0.955)
	50%	-0.059(0.939)	0.047(0.974)
200	25%	-0.025(0.955)	0.004(0.941)
	50%	0.014(0.945)	0.020(0.935)

%, censoring percentage

Table 2: Regression analysis of brain cancer trial accounting for treatment effectiveness lag time

Model			PHM			ADHM		
	Parm	Cov	Est	SE	95%CI	Est	SE	95%CI
Hazard λ	β	Treatment	-0.227	0.140	(-0.501,0.047)	-0.005	0.003	(-0.001,0.011)
	γ	Age	0.221	0.054	(0.115,0.327)	0.005	0.001	(0.003,0.007)
Response p	α	Treatment	-	-	-	-	-	-
		Age	-	-	-	-	-	-
F_0	μ	-	-	-	-	-	-	-

Model			ADHM-L		
	Parm	Cov	Est	SE	95%CI
Hazard λ	β	Treatment	-0.014	0.009	(-0.032,-0.028)
	γ	Age	0.006	0.001	(0.004,0.008)
Response p	α	Treatment	-0.002	3.063	(-6.005,6.001)
		Age	-0.277	0.509	(-1.275,0.721)
F_0	μ		2.502	1.037	(0.469,4.535)

PHM, proportional hazards model; ADHM, additive hazards model; ADHM-L, additive hazards model with latent treatment effectiveness lag time; Parm, parameter; Cov, covariate; Est, estimate; SE, estimated standard error; 95%CI, 95% confidence interval.

Figure 1: Lag functions of $H(t;\beta_0,\phi_0)$ with truncation time $u_0 = \infty$. $p(\alpha)$ is proportion of short-term lag times. τ is parameter for exponential distribution of short-term lag times.

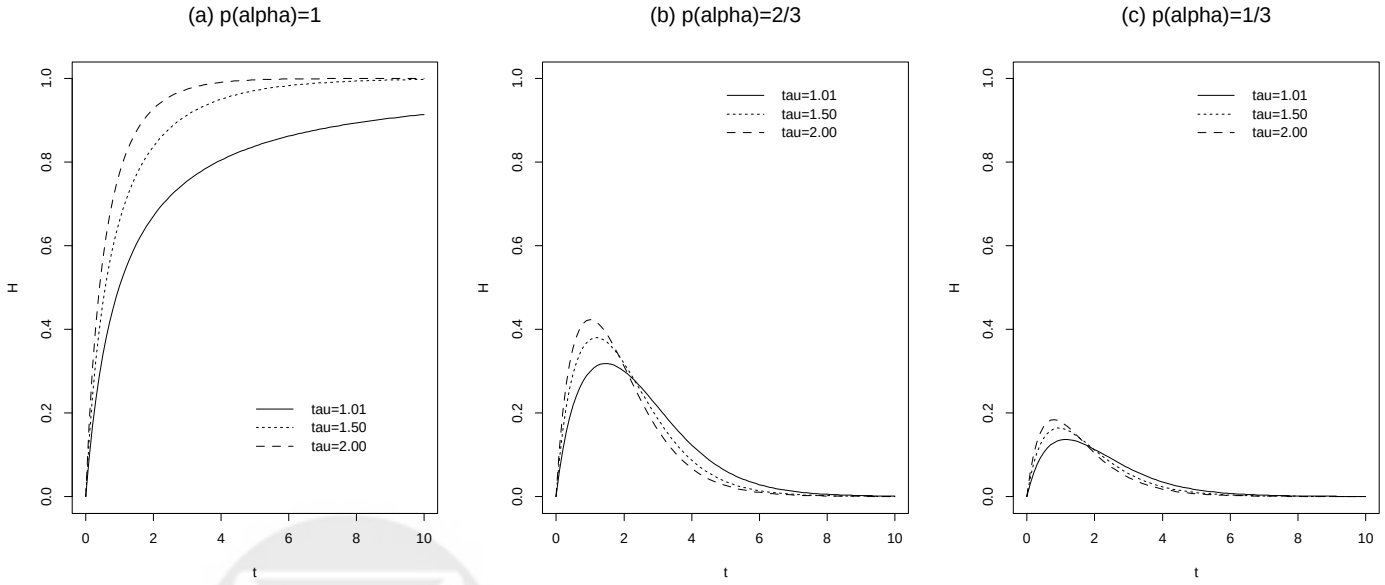

Figure 2: Lag functions of $H(t;\beta_0,\phi_0)$ with truncation time $u_0 = 5$. $p(\alpha)$ is proportion of short-term lag times. τ is parameter for exponential distribution of short-term lag times.